Catching Click Fraud

Vacha Dave, Saikat Guha
Microsoft Research India
Internet Advertising Today

Online advertising is a 37 billion dollar industry
15% year-on-year growth

Publishers can monetize traffic

Blogs, News sites, Syndicated search engines
Revenue for content development

Pay-per-click advertising

Advertisers pay per-click to ad networks

Publishers make a 70% cut on each click on their site

*Based on Interactive Advertising Bureau Report, a consortium of Online Ad Networks
Click-Fraud vs. Click-Spam
Click-Spam Problem

No ground truth
Almost impossible to know if particular click is genuine
Need to guess the intent of user

No reliable information from industry
Ad networks: security by obscurity
Third parties: proprietary heuristics contested by ad networks

No research in academia
Researchers lack real attack data
Research Contributions

First method to **independently estimate** click-spam

As an advertiser, research, anybody.
For specific keywords. No heuristics.

Test across ten ad networks

Search, contextual, social and mobile ad networks
Demonstrate and quantify that click-spam is an open (research) problem
Especially for Mobile and Social ad networks

Discover **six classes of sophisticated attacks**

Why simple heuristics don’t work

**Release data for researchers**
Active Measurements

Lots of clicks

<table>
<thead>
<tr>
<th>Celebrity Cruise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be Recognized, Celebrated, And Indulged Aboard Celebrity Cruises</td>
</tr>
</tbody>
</table>

How many 🙁❓?

\[ x = \frac{g_d (l_i - l'_i \frac{d}{dt})}{n_d g_i} \]

Lots of visits

Random sample

<table>
<thead>
<tr>
<th>Celebrity Cruise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be Recognized, Celebrated, And Indulged Aboard Celebrity Cruises</td>
</tr>
</tbody>
</table>

Black Box

Fewer visits

Collect Clean Data → Detect Click-Fraud

BingAds Log Data → Gain Insights

Catch Click-Fraud

\[ x = \frac{g_d (l_i - l'_i \frac{d}{dt})}{n_d g_i} \]
Hurdles, Bluff ads

**Celebrity Cruise**
Be Recognized, Celebrated, And Indulged Aboard Celebrity Cruises
cruisewithceleb.com

Lots of clicks

**Massive smile Literature**
Cream Fix Gutter Bad Keys
cruisewithceleb.com

Bluff

“Loading…”, or “Click to continue”

Few clicks. Spammers and (few) curious good users.

Fewer spammers. Slightly fewer good users.
Testing Ad Networks

Sign up as advertisers for ten ad networks
Search, Contextual, Mobile and Social
Google, Bing, AdMob, InMobi, Facebook and others

240 Ads
Keywords: Celebrity, Yoga, Lawnmower
Hurdles: Click to continue, 5 sec wait

50,000 Clicks
30,000 bait ad clicks

Cost: $1500
Validation and Results

Ad Network’s Estimate

Valid Traffic Fraction (Normalized)

Ad Networks

- Our Estimate

celebrity

Ad Networks

Collect Clean Data

Gain Insights

Catch Click-Fraud

BingAds Log Data
Where from?

Analyze bait ad clicks
Publishers: Strong motive
Instead of clicks/users

Manual Investigation

Challenge: Scale
3000+ publishers, 30,000 Clicks

Identical sites!

Cluster on cosine similarity
Feature vector
WHOIS, IP Address/Subnet, HTTP parameters
Arbitrage
Adult
Parked
Malware
Case Study 1 – Malware

Responsible Malware: TDL4

Validation: Run malware in VM

Browser specific filtering doesn’t work
Can intercept and redirect all browser requests

Timing analysis doesn’t work
Mimics real user behavior

Threshold based filtering doesn’t work
Only 1 click per IP address per day

Signature
Large number of users, Small number of clicks (per day), Sustained clicks over multiple days
Massive Low Intensity Attacks

- bing.com
- coolsearchnow.com
- meetgirlstonight.net
- jjbargains.com

# Days Clicked in a Week

# Users
Case Study 2 – Arbitrage

Discussion Forum
Low page-rank, not malicious, no ads
No typos, no dates
Periodically reset

Cheap Ads!
Page now full of (expensive) ads
Confused user clicks

Work From Home-Now Hiring
3 Positions Available - $17-21/Hour Based On Experience. Apply Now!
DoTellAll.com/Apply
Ongoing Tech Transfer and Research

1. General Insight
2. Adaptive adversary
3. Algorithm public

Collect Clean Data

Detect Click-Fraud

Catch Click-Fraud

Tech Transfer

Published Paper
Research Contributions

First method to independently estimate click-spam

As an advertiser, research, anybody.
For specific keywords. No heuristics.

Test across ten ad networks

Search, contextual, social and mobile ad networks
Demonstrate and quantify that click-spam is an open (research) problem
Especially for Mobile and Social ad networks

Discover six classes of sophisticated attacks

Why simple heuristics don’t work

Release data for researchers